# Image Saliency by Isocentric Curvedness and Color

Roberto Valenti Nicu Sebe Theo Gevers
Intelligent Systems Lab Amsterdam
University of Amsterdam, Amsterdam, The Netherlands
{r.valenti,n.sebe,th.gevers}@uva.nl

## Abstract

*In this paper we propose a novel computational method to infer visual saliency in images. The method is based on the idea that salient objects should have local characteristics that are different than the rest of the scene, being edges, color or shape. By using a novel operator, these characteristics are combined to infer global information. The obtained information is used as a weighting for the output of a segmentation algorithm so that the salient object in the scene can easily be distinguished from the background.*

*The proposed approach is fast and it does not require any learning. The experimentation shows that the system can enhance interesting objects in images and it is able to correctly locate the same object annotated by humans with an F-measure of* $85.61\%$ *when the object size is known, and* $79.19\%$ *when the object size is unknown, improving the state of the art performance on a public dataset.*

## 1. Introduction

Visual saliency is a very important part of our vision: it is the mechanism that helps in handling the overload of information that is in our visual field by filtering out the redundant information. It can be considered as a measure determining to what extent an image area will attract an eye fixation. Unfortunately, little is known about the mechanism that leads to the selection of the most interesting (salient) object in the scene such as a landmark, an obstacle, a prey, a predator, food, mates etc. It is believed that interesting objects on the visual field have specific visual properties that makes them different than their surroundings. Therefore, in our definition of visual saliency, no prior knowledge or higher-level information about objects is taken into account.

In this paper, we are interested in the computational simulation of this mechanism, which can be used in various computer vision scenarios, spanning from algorithm optimization (less computation spent on uninteresting areas in the image) to image compression, image matching, content-based retrieval, etc.

General *context-free* (*i.e.* without prior knowledge about the scene) salient point detection algorithms aim to find distinctive local events in images by focusing on the detection of corners, edges [3, 27, 18, 12] and symmetry [20, 13, 4]. These methods are very useful to find locally salient points, however globally salient regions are usually computed by partitioning the images into cells and by counting the number of salient descriptors which fall into them. The above techniques lack the ability to infer the location of global structures as an agreement of multiple local evidences. Therefore, to infer global salient regions, in this paper, we propose a framework that combines isophotes properties (Section 2.1) with image curvature (Section 2.2) and color edges information (Section 2.3). The contributions are the following:

- Instead of using the widely adopted edge information, we propose to use the gradient slope information to detect salient regions in images.
- We use an isophote symmetry framework to map local evidence close to the centers of image structures.
- We provide an enabling technology for smarter, saliency aware, segmentation algorithms.
- We solve the problem of defining the size of unknown interesting objects by segmentation and subwindow search.

## 2. The Saliency Framework

By analyzing human vision and cognition, it has been observed that visual fixations tend to concentrate on corners, edges, along lines of symmetry and distinctive colors. Therefore, previously proposed saliency frameworks in the literature [11, 5, 1, 8, 17, 21, 19] often use a combination of intensity, edge orientation and color information to generate saliency maps. However, most interest detectors focus on the shape-saliency of the local neighborhood, or point out that salient points are "interesting" in relation to their direct surroundings. Hence, salient features are generally determined from the local differential structure of images.

In this paper, the goal is to go from the local structures to more global structures. To achieve this, based on the observation that the isophote framework (previously proposed in [24] for eye detection) can be generalized to extract



2009 IEEE 12th International Conference on Computer Vision (ICCV)
978-1-4244-4419-9/09/$25.00 ©2009 IEEE

generic structures in images, we propose a new isophote-based framework which uses additional color edges and curvature information. As with many other methods proposed in the literature, our approach is inspired by the feature integration theory [22]. Therefore, we compute different salient features which are later combined and integrated into a final saliency map. In the next sections, the principles of each of the used salient features and how they are combined are described.

## 2.1. Isocentric Saliency

Isophotes are lines connecting points of equal intensity (curves obtained by slicing the intensity landscape). Since isophotes do not intersect each other, an image can be fully described by its isophotes both on its edges and on smooth surfaces [10]. Furthermore, the shape of each isophote is independent of changes in the contrast and brightness of an image. Due to these properties, isophotes have been successfully used as features in object detection and image segmentation [10, 6]. To formulate the concept of isophote, a local coordinate system is defined at every point in the image, which points in the direction of gradient. Let the gauge coordinate frame be $\{v, w\}$, the frame vectors can be defined as

$$\hat{w} = \frac{\{L_x, L_y\}}{\sqrt{L_x^2 + L_y^2}}, \qquad \hat{v} =\perp \hat{w},$$

where $L_x$ and $L_y$ stand for the first-order derivatives of the luminance function $L(x, y)$ in the $x$ and $y$ dimensions, respectively. Since by definition there is no change in intensity on an isophote, the derivative along $v$ is 0, whereas the derivative along $w$ is the gradient itself. Thus, an isophote is defined as $L(v, w(v)) = constant$.

At each point of the image, we are interested in the displacement of the center of the osculating circle to the isophote, which is assumed to be not far away from the center of the structure to which the isophote belongs. Knowing that an isophote is a curvilinear shape, the isophote curvature, $\kappa$, is computed as the rate of change, $w''$, of the tangent vector, $w'$. In Cartesian coordinates, this is expressed as:

$$\kappa = -\frac{L_{vv}}{L_w} = -\frac{L_y^2 L_{xx} - 2L_x L_{xy} L_y + L_x^2 L_{yy}}{(L_x^2 + L_y^2)^{3/2}}.$$

The magnitude of the vector (radius) is simply found as the reciprocal of the above term. The information about the orientation is obtained from the gradient, but its direction indicates the highest change in luminance. The duality of the isophote curvature is then used in disambiguating the direction of the vector: since the sign of the isophote curvature depends on the intensity on the outer side of the curve, the gradient is simply multiplied by the inverse of the isophote curvature. Since the gradient is $\frac{\{L_x, L_y\}}{L_w}$ , the displacement coordinates $D(x, y)$ to the estimated center are obtained by

$$\begin{aligned} D(x,y) &= \frac{\{L_x, L_y\}}{L_w}\left(-\frac{L_w}{L_{vv}}\right) = -\frac{\{L_x, L_y\}}{L_{vv}} \\ &= -\frac{\{L_x, L_y\}(L_x^2 + L_y^2)}{L_y^2 L_{xx} - 2L_x L_{xy} L_y + L_x^2 L_{yy}}. \end{aligned} \quad (1)$$

In this manner every pixel in the image gives an estimate of the potential structure it belongs to. In order to collect and reinforce this information and to deduce the location of the objects, $D(x, y)$'s are mapped into an accumulator, which is in turn convolved with a Gaussian kernel so that each cluster of votes will form a single estimate. This clustering of votes in the accumulator gives an indication of where the centers of interesting or structured objects are in the image (isocenters). By applying this framework to natural images, many votes can be affected by noise or are generated by uninteresting cluttered parts of the image. To reduce this effect, each vote is weighted according to its local importance, defined as the amount of image curvature (Section 2.2) and color edges (Section 2.3).

## 2.2. Curvature Saliency

A number of approaches use edge information to detect saliency [27, 18, 12]. The amount of information contained in edges is limited if compared to the rest of the image. Instead of using the peaks of the gradient landscape, we propose to use the slope information around them. To this end, an image operator that indicates how much a region deviates from flatness is needed. This operator is the curvedness [7], defined as

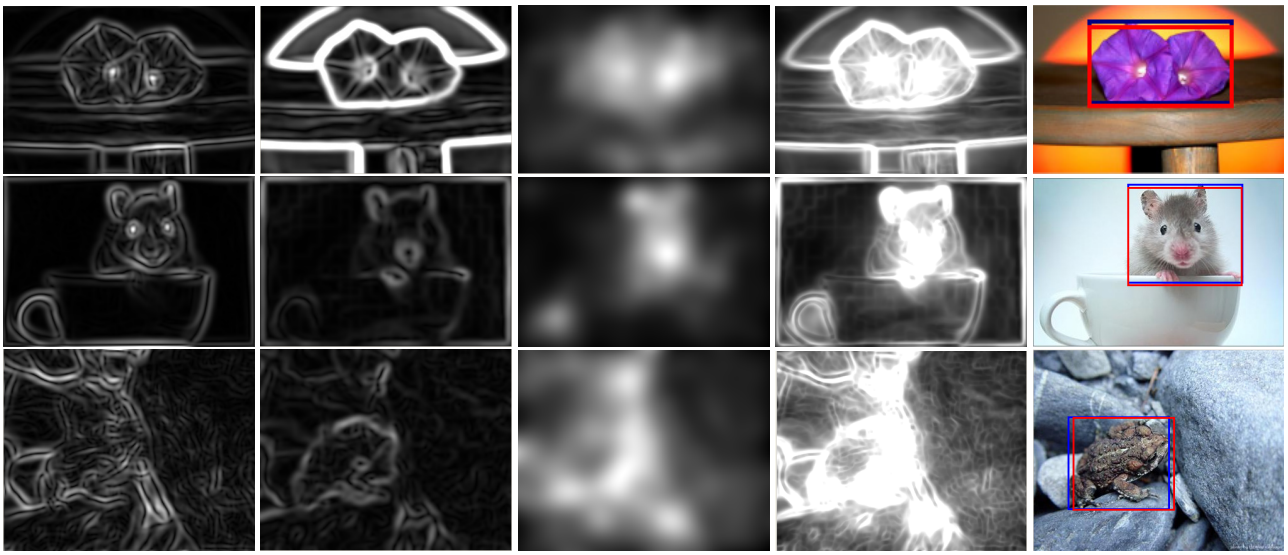
$$curvedness = \sqrt{L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2}.$$

The curvedness can be considered as a rotational invariant gradient operator, which measures the degree of regional curvature. Since areas close to edges will have a high slope and since isophotes are slices of the intensity landscape, there is a direct relation between the curvedness and the density of isophotes. Hence isophotes with higher curvedness are more appropriate for our goal of mapping from local structures to global structures, as they are likely to follow object boundaries and thus belong to the same shape. An example of the effect obtained by applying the curvedness to natural images can be seen in Figure 1(a).

## 2.3. Color Boosting Saliency

While determining salient image features, the distinctiveness of the local color information is commonly ignored. To fully exploit the information coming from the color channels, both shape and color distinctiveness should be taken into account.

The method proposed by van de Weijer *et al*. [25] is based on the analysis of the statistics of color image derivatives and uses information theory to boost the color information content of a color image. Since color boosting is



(a) (b) (c) (d) (e)

Figure 1. An example of the conspicuity maps and their combination: (a) Curvedness, (b) Color boosting, (c) Isocenters clustering, (d) Combined saliency map, (e) Area with highest energy in the saliency map (red: detection, blue: ground truth).

derivative based and its outcome is enhanced color edges, the method can easily be integrated in our framework. According to information theory, rare events are more important than normal events. Therefore, the quantity of information $I$ of a descriptor $v$ with probability $p(v)$ can be defined as

$$I(v) = -\log(p(v)).$$

In order to allow rare color derivatives to have equal values, image derivatives are mapped to a new space using a color saliency boosting function $g$. The function $g$ can be created by analyzing the distribution of image derivatives. In fact, it can be derived that the shape of the distribution is quite similar to an ellipse that can be described by the covariance matrix $M$. This matrix can be decomposed into an eigenvector matrix $U$ and an eigenvalue matrix $V$. The color boosting function $g$ will be the transformation with which the ellipses are transformed into spheres: $g(L_x) = V^{-1}U^T L_x$, where the eigenvectors $U$ are restricted to be equal to the opponent color space, and $V = diag(0.65, 0.3, 0.1)$ as was found in the distribution of the data in the Corel dataset [25]. After the mapping, the norm of the image derivatives is proportional to the information content they hold. An example of the effect obtained by applying this operator to natural images can be seen in Figure 1(b).

## 3. Building the Saliency Map

In this section, the previously described saliency features are combined into a saliency framework. Since all the features sections make use of image derivatives, their computation can be re-used in order to lower the computational costs of the final system. Furthermore, the three features were selected as both curvedness and color boosting enhance edges, and isophotes are denser around edges. Therefore, the saliency features can be nicely coupled together to generate three different conspicuity maps (Figure 1(a), (b) and (c)): At first, the maps obtained by the curvedness and the color boosting are normalized to a fixed range $[0, 1]$ so that they can easily be combined. The linear combination of these maps is then used as weighting for the votes obtained from Eq. 1 to create a good isocenter clustering of the most salient objects in the scene. In this way, the energy of local important structures can contribute to find the location of global important structures. An example of isocenter clustering is given in Figure 1(c), obtained by weighting the votes for the isocenters using the curvature and color boosting conspicuity maps in Figure 1(a) and (b). The main idea is that if a region of the image is relevant according to multiple conspicuity maps, then it should be salient, therefore the normalized conspicuity maps are linearly combined into the final saliency map (Figure 1(d)). However, multiple objects or components could be present in an image and hence receive higher saliency energy from the conspicuity maps. For instance, in the example on the second line of Figure 1, the mouse and the handle of the cup are receiving the most of the energy, but what is the real salient object? The full mouse, its face, the cup, the handle or all of them together? This question raises the problem of scale and size of the object that we are looking for. Depending on the application, the size of the object might be known (*e.g.* the size of the silhouette of a person seen from a specific security camera, the size of the object on which we are performing visual inspection for quality control *etc.*). In the experimental sec-

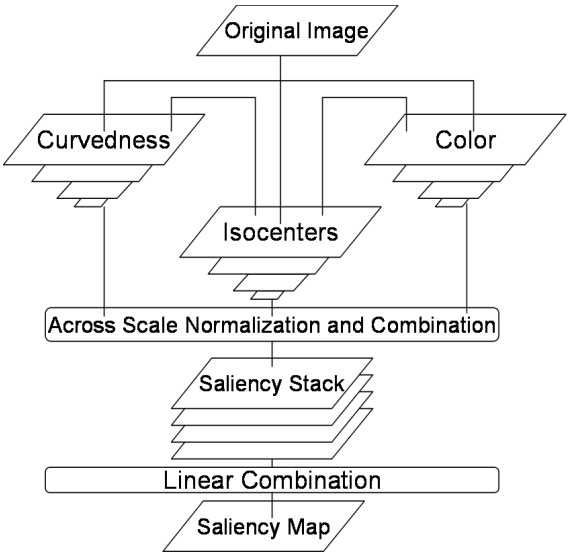



Figure 2. The proposed saliency scale space framework

tion we will show that, if the size is known, the saliency map obtained with the described procedure is already enough to obtain a good location estimate for the salient object. This is shown in Figure 1(e), where the red box represents the area with the maximum energy (the salient region in the image), and the blue box the ground truth annotation. On the other hand, if the size of the object is unknown, additional information about the persistence of the object in scale space and information about its boundaries are required. These topics are discussed in the following sections.

### 3.1. Scale Space

The scale selection is an important problem that must be considered when defining what a salient object is. The scale problem is commonly solved by exhaustively searching for the scale value that obtains the best overall results on a homogeneous dataset. Given the heterogeneity of the size of images and the depicted objects, we want to gain scale independence in order to avoid adjustments of the parameters for different situations.

To increase robustness and accuracy, a scale space framework is used to select the results of the conspicuity maps that are stable across multiple scales. To this end, a Gaussian pyramid is constructed from the original color image. The image is convolved with different Gaussians so that they are separated by a constant factor in scale space. In order to save computation, the image is downsampled into octaves. In each octave the conspicuity maps are calculated at different intervals: for each of the image in the pyramid, the proposed method is applied by using the appropriate $sigma$ as a parameter for the size of the kernel used to calculate image derivatives. This procedure results in two saliency pyramids (Figure 2), one retaining the color saliency and the other the curvature saliency. These two pyramids are then combined together with isophote information to form a third saliency pyramid, containing isocentric saliency. The responses in each of the three saliency pyramids are combined linearly, and then scaled to the original image size to obtain a scale space saliency stack. Every element of the saliency stack is normalized and therefore considered equally important, hence they are simply accumulated into a single, final saliency map. The areas with the highest energy in the resulting saliency map will represent the most scale invariant interesting object, which we will consider to be the object of interest in the image.

### 3.2. Graph Cut Segmentation

Although the obtained saliency map has most of its energy at the center of image structures, the successful localization of the most salient object can only be achieved by analyzing its edges. In fact, since curvedness and color boosting are combined together with isocentric saliency in the final saliency map, a great part of the energy in it will still lie around edges. To distribute the energy from the center and the edges of the salient structure to connected regions, a fast and reliable segmentation algorithm is required. The method proposed in [2] addresses the problem of segmenting an image into regions by using a graph-based representation of the image. The authors propose an efficient segmentation method and show that it produces segmentations that satisfy global properties. This method was chosen in this paper because its computational complexity is nearly linear in the number of graph edges and it is also fast in practice. Furthermore, it can preserve the details in low-variability image regions while ignoring them in high-variability regions.

As shown in [23], it is possible to extract a salient object in an image by means of a segmentation algorithm and eye fixations. Since the saliency map obtained by our system can be considered as a distribution of potential fixation points, it can be used to enhance the segmentation algorithms to extract connected salient components. The graph cut segmentation results for a set of images are shown in the second row of Figure 3. For each of the segmented components, the average energy covered in the saliency map is computed. Therefore, if the component has higher energy, it will be highlighted in the saliency weighted segmentation.

The third row of Figure 3 shows the effect of weighting the segmentation components on second row of Figure 3 by the saliency map in the first row of Figure 3. Note that, in this case, the brightness indicates the level of saliency of the region (brighter = more salient) and that, if the results are thresholded, it is possible to obtain a binary map of segmented salient regions. This opens the possibility for saliency based segmentation algorithms that would join segmented components based on their saliency other than their

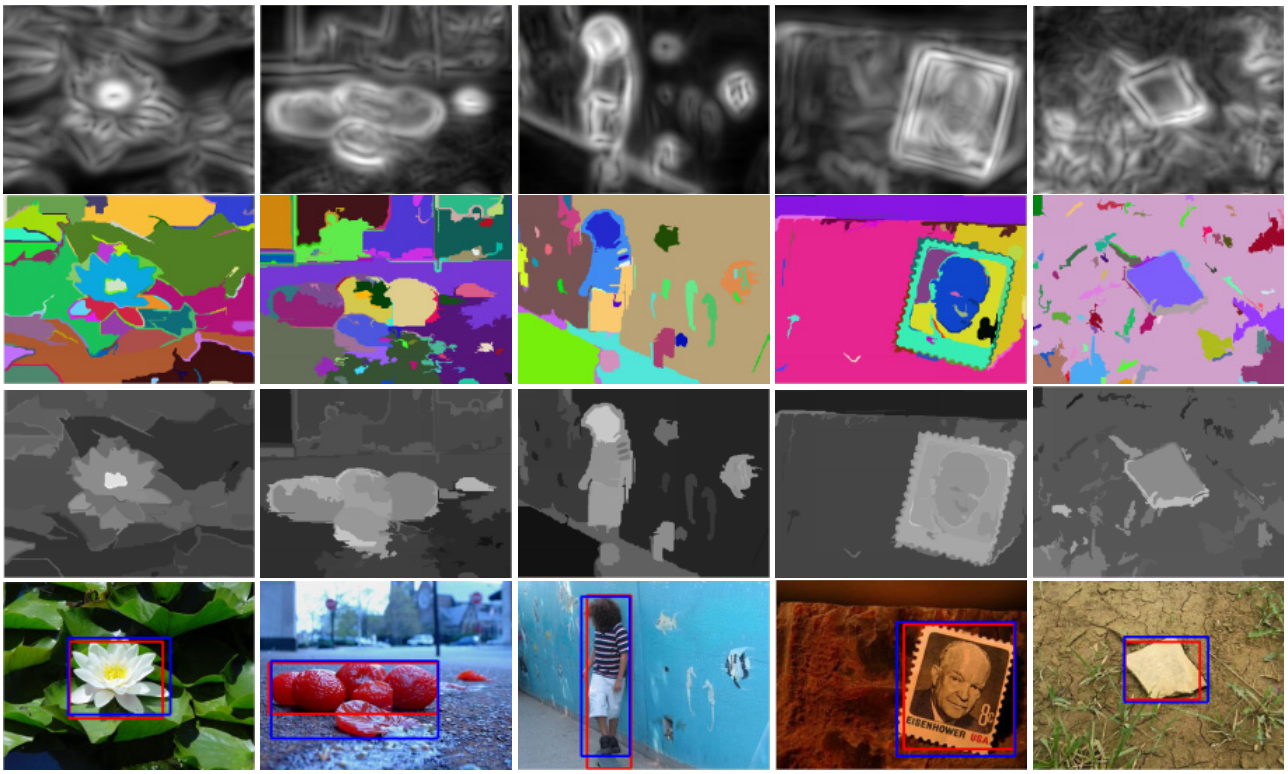


Figure 3. An example of the obtained results. From top to bottom: Saliency map, graph-cut segmentation, segmentation weighted by saliency, ESS result (red: detection, blue: ground truth).

similarity.

## 4. Experiments

In this section, the accuracy of the proposed algorithm and of its components is extensively evaluated on a public dataset.

### 4.1. Dataset and Measures

The used saliency dataset is the one reported in [11]. The dataset contains 5000 high quality images, each of them hand labeled by 9 users requested to draw a bounding box around the most salient object (according to their understanding of saliency). The provided annotations are used to create a saliency map $S = \{s_x | s_x \in [0, 1]\}$ as follows:

$$s_x = \frac{1}{N} \sum_{n=1}^{N} a_x^n$$

where $N$ is the number of users and $a_x^n$ are the pixels annotated by user $n$. In this way, the annotations are combined into an average saliency map. In order to create a binary ground truth saliency map, only the bounding box of the area annotated by more than four users is kept as annotation $s_x$ of the most salient object in the scene. Given the ground truth annotation $s_x$ and the obtained detection $d_x$ of the salient region in an image, the precision, recall, and F-measure are calculated. The precision and recall measures are defined as:

$$Precision = \frac{\sum_x s_x d_x}{\sum_x d_x} \qquad Recall = \frac{\sum_x s_x d_x}{\sum_x s_x}.$$

The F-measure is the weighted harmonic mean of precision and recall, therefore is an overall performance measure. It is defined as

$$F\text{-}measure = \frac{(1+\alpha) \times Precision \times Recall}{\alpha \times Precision + Recall},$$

where $\alpha$ is set to 0.5 as in [11] and [16]. All measures are then averaged over all the 5000 images in the dataset to give overall figures.

### 4.2. Methodology

The task of determining the location and size of an unknown object in an image is very difficult. The proposed system is tested against two scenarios: one in which the size of the interesting object is known and one where no assumptions on the size are made.

#### 4.2.1 Sliding Window

The purpose of this test is to verify whether or not the location of an interesting object can be retrieved if its scale

| Method | Size Known | | | Size Unknown | | |
|---|---|---|---|---|---|---|
| | Precision | Recall | F-measure | Precision | Recall | F-measure |
| Curvedness | 77.55% | 77.11% | 77.40% | 72.47% | 50.74% | 49.95% |
| Isocenters | 79.95% | 79.49% | 79.79% | 84.23% | 66.39% | 72.44% |
| Color | 80.91% | 80.45% | 80.75% | 81.63% | 37.29% | 44.41% |
| Curvedness + Color | 83.79% | 83.31% | 83.63% | 71.50% | 71.73% | 67.29% |
| All | 85.77% | 85.28% | 85.61% | 84.91% | 76.19% | 79.19% |

Table 1. Contribution of each of the used features and their combination, when the size of the object is known or unknown.

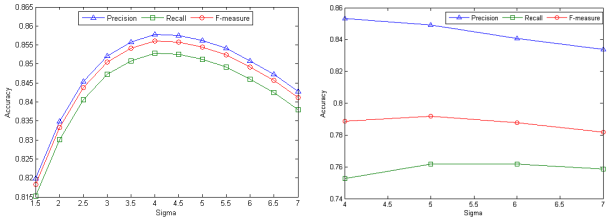
is known. Therefore, in this scenario, the size of the object is known, and it corresponds to the size obtained from the ground truth. The location of the object in the image, however, is unknown. With the help of integral images (as defined in [26]), the saliency map is exhaustively searched for the region $d_x$ which obtains the highest energy by sliding the ground-truth window over it.

#### 4.2.2 Efficient Subwindow Search

In this scenario, both the size and the location of the relevant object in the image are unknown. To solve this problem, an exhaustive search of all possible subwindows in the saliency map could be performed, in order to retain the one which covers the most energy. However, this would be computationally unfeasible. The Efficient Subwindow Search (EES) is an algorithm which replaces sliding windows approaches to object localization by a branch-and-bound search [9]. It is a simple yet powerful scheme that can extend many existing recognition methods to also perform localization of object bounding boxes. This is achieved by maximizing the classification score over all possible subwindows in the image. The authors show that it is possible to efficiently solve a generalized maximum subwindow problem in many situations. However, even if an efficient search of the best subwindow could be performed, not knowing the size of the object will result in many subwindows with high energy (as in the mouse example in Section 3). To obviate this problem, the ESS algorithm is applied on the integral image of the saliency weighted segmentation (third row of Figure 3), obtained as described in Section 3.2.

### 4.3. Evaluation

In order to estimate the partial contribution of each of the conspicuity maps to the final saliency map, they are evaluated independently. The sliding window methodology is used to obtain results that are independent of segmentation. Therefore, only the discriminant saliency power of the features is evaluated. By simply using the curvedness, the method can already achieve a good estimate of the salient object in the image (F-measure $77.40\%$). However, curvedness alone fails in cluttered scenes, as the number of edges will distract the energy from the salient object. The plain isocenters clustering (without weighting) obtains better performance (F-measure $79.79\%$), similar to color




(a) (b)

Figure 4. Effect of changing the standard deviation of the Gaussian kernel when the object size is known (a) and unknown (b).

boosting alone (F-measure $80.75\%$). The linear combination of color boosting and curvedness provides an improved result over the two features considered independently (F-measure $83.63\%$). This indicates that the two features are somewhat complementary (one succeeds when the other fails). Finally, the proposed combination of all the features achieves the best result (F-measure $85.61\%$).

In the second scenario, the used edge features are expected to fail as they do not contribute to the center of the components. In fact, curvedness and color boosting achieve an F-measure of $49.95\%$ and $44.41\%$, while their combination only improves this figure to $67.29\%$. However, given its capability to distribute the boundary energy to the center of image structures, the isocenter saliency alone has an F-measure of $72.44\%$. The combination of all the features achieves an F-measure of $79.19\%$. A summary of the obtained precision, recall and F-measure accuracy for each of the features in both scenarios is shown in Table 1.

The graphs in Figure 4 show how the accuracy changes with respect to the used standard deviation of the Gaussian kernel ($sigma$) in both scenarios. In the first scenario the $sigma$ parameter can be fine tuned to obtain the best results. Note that, since the size of the object corresponds to the size in the ground truth, there is a relation between precision, recall and F-measure and they are therefore very similar. In the second scenario, when using the ESS search over the saliency weighted segmentation, changing the parameter has virtually no effect on the accuracy of the system as it has the sole effect of slightly modify the energy in each the segmentation components.

We compared our results with the ones obtained by other methods in the literature: The method from Ma *et al.* [14] uses fuzzy growing, the framework proposed by Itti *et*

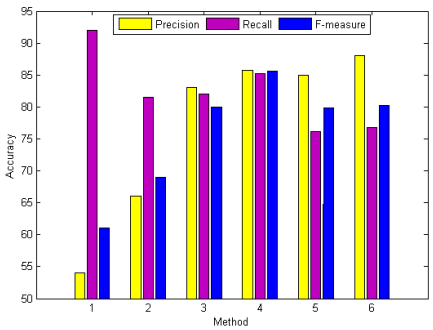



Figure 5. A comparison with other methods: 1) Ma *et al*. [14] 2) Itti *et al*. [5] 3) Liu *et al*. [11] 4) Our method on the first scenario (size known) 5) Our method on the second scenario (size unknown) 6) Worst human annotation.

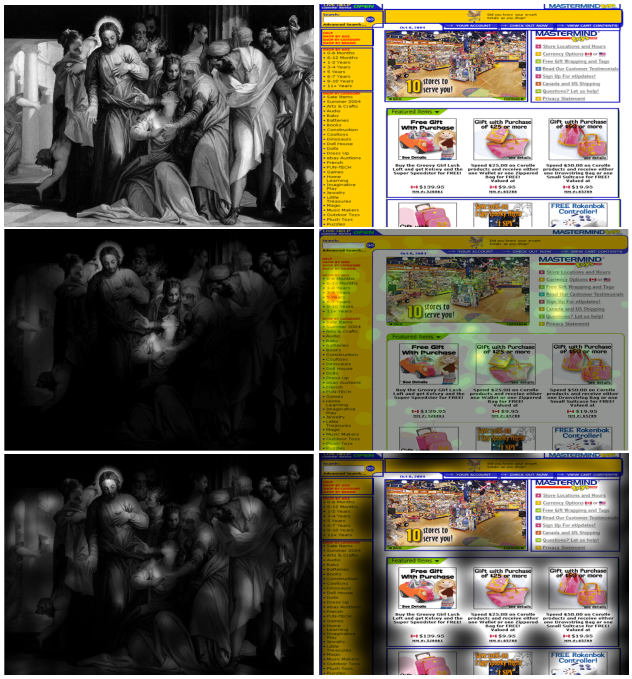
*al*. [5] uses multiscale color, intensity and orientation features, and the method from Liu *et al*. [11] uses multi-scale contrast, center-surround histogram and color spatial distribution, combined by a learned Conditional Random Field. To compare the quality of the obtained result with respect to a human performance, we computed the worst performing human annotation with respect to the ground truth annotation (obtained by agreement of at least four subjects). This corresponds to a precision of $87.99\%$, a recall of $76.74\%$ and an F-measure of $80.29\%$.

A summary of the precision, recall, and F-measure achieved by the cited methods is displayed in Figure 5. Note that our first scenario (column 4) has prior knowledge about the size of the object and is therefore not directly comparable with the other methods. However, without using any prior knowledge (column 5), our method outperforms the classical approaches [14, 5] on the same dataset, while achieving comparable results with the state of the art method [11] without requiring any learning.

Furthermore, as already discussed in [11], the precision measure is much more important in saliency than the recall (*e.g*. if all the image is selected as the salient region, the recall is $100\%$). In our case, we obtain the highest precision when compared to the state of the art methods, while achieving the same F-measure as the best computational methods and as the worst human annotation.

### 4.4. Visually Salient vs. Semantically Salient

In order to discriminate if an object is visually or semantically interesting, we illustrate in Figure 6 a qualitative comparison between the obtained saliency map and heat maps obtained by analyzing eye fixations on the same images. By analyzing the painting example in Figure 6(a) it is clear that there is a similarity between the eye fixations (second row) and the detected salient regions (third row). It can be seen that, even if they are equally visually salient, the subject appears to mainly focus on faces as they are more



(a) (b)

Figure 6. A comparison with eye fixations. From top to bottom: Original image, recorded eye fixations, saliency map obtained by our method superimposed to the original image.

semantically salient and less on lower areas, which hold less semantic information (like knees). The same reasoning can be done for the website example in Figure 6(b): while every line of the navigation menu is equally salient, the subject focuses only on the top entries. Also, while the items in the middle of the page have similar visual saliency, the user seems to focus only on few of them. This is a clear difference between visual saliency and higher levels of reasoning (*e.g*. knowledge and interest), which can be used to understand if an object is semantically interesting versus visually interesting. It can be seen, however, that eye fixations are always directly related with salient regions in the image. Therefore, if eye fixation information is available, our method could be used to differentiate between salient regions and the subject's interest. This information can be very valuable as it could be used in a multitude in applications (*e.g*. to tailor user interfaces or commercial ads in minimizing the possible elements of distraction in the visual field).

### 4.5. Discussion

By analyzing our results, we found that the main reason for the low recall lies on the manner that the dataset is annotated: by considering the fruit example in the second column of Figure 3, it can be seen that the detected region is smaller than the annotated one. However, the saliency weighted segmentation of the salient object is nearly op-

timal. By including this last part of the object in the detected subwindow, a big part of the background would get included as well, lowering the overall energy covered by the subwindow, which in turn would be discarded by the ESS algorithm. The same happens in many of the images in the dataset, explaining our low recall: appendices of object are often not considered as they would decrease the overall energy in the detected subwindow.

As the proposed system focuses on images, we are aware that it is not suitable for locating the salient object in all situations, especially involving changes and movements (as in [15]). Given the flexibility of the framework, the conspicuity maps from other saliency operators can easily be added to the system in order to cover additional saliency cues. However, this will probably require some learning to correctly integrate the different conspicuity maps, and thereby reduce the attractiveness of our method as, contrary to other systems, the actual creation of the saliency map is computationally fast (only a combination of few image derivatives is needed) and does not require any training.

## 5. Conclusions

In this paper, we have presented a computational bottom-up model to detect visual saliency in common images. The method is based on the assumption that interesting objects on the visual field have specific structural properties that makes them different than their surroundings, and that they can be used to infer global important structures in the image.

The system performs well as it is able to correctly locate or give maximum energy to the same object annotated by humans with an F-measure of $85.61\%$ if the size of the object is known. If the size of the object is unknown, our method is used to enhance a segmentation algorithm. An efficient subwindow search on the saliency weighted segmentation shows that the algorithm can correctly locate an interesting object with an F-measure of $79.19\%$, while keeping a high precision. The obtained results are very promising as they match the worst human annotation. Furthermore, since no learning is required but only calculation of image derivatives, the system is fast and it can be used as a preprocessing step in many other applications.

## References


[1] L. Elazary and L. Itti. Interesting objects are visually salient. *Journal of Vision*, 8(3):1–15, 3 2008.

[2] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 59(2), 2004.

[3] C. Harris and M. Stephens. A combined corner and edge detection. In *Alvey Vision Conf.*, pages 147–151, 1988.

[4] G. Heidemann. Focus-of-attention from local color symmetries. *PAMI*, 26(7):817–830, 2004.

[5] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *PAMI*, 20(11):1254–1259, 1998.

[6] C. Kervrann, M. Hoebeke, and A. Trubuil. Isophotes selection and reaction-diffusion model for object boundaries estimation. *IJCV*, 50:63–94, 2002.

[7] J. Koenderink and A. van Doorn. Surface shape and curvature scales. *Im. and Vis. Comp.*, pages 557–565, 1992.

[8] L. Kovacs and T. Sziranyi. Focus area extraction by blind deconvolution for defining regions of interest. *PAMI*, 29(6):1080–1085, 2007.

[9] C. Lampert, M. Blaschko, and T. Hofmann. Beyond sliding windows: Object localization by efficient subwindow search. *CVPR*, 2008.

[10] J. Lichtenauer, E. Hendriks, and M. Reinders. Isophote properties as features for object detection. In *CVPR*, 2005.

[11] T. Liu, J. Sun, N. N. Zheng, X. Tang, and H. Y. Shum. Learning to detect a salient object. *CVPR*, 2007.

[12] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 20:91–110, 2003.

[13] G. Loy and A. Zelinsky. Fast radial symmetry for detecting points of interest. *PAMI*, 25(8):959–973, 2003.

[14] Y. F. Ma and H. J. Zhang. Contrast-based image attention analysis by using fuzzy growing. In *ACM MM*, 2003.

[15] S. Marat, T. Phuoc, L. Granjon, N. Guyader, D. Pellerin, and A. Guerin. Modelling spatio-temporal saliency to predict gaze direction for short videos. *IJCV*, 82:231–243, 2009.

[16] D. R. Martin, C. C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *PAMI*, 26(5):530–549, 2004.

[17] O. L. Meur, P. L. Callet, D. Barba, and D. Thoreau. A coherent computational approach to model bottom-up visual attention. *PAMI*, 28(5):802–817, 2006.

[18] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *PAMI*, 19(5):530–534, 1997.

[19] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Robust object recognition with cortex-like mechanisms. *PAMI*, 29(3):411–426, Feb 2007.

[20] H. Tek and B.Kimia. Symmetry maps of free-form curve segments via wave propagation. *IJCV*, 54:35–81, 2003.

[21] A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol Rev*, 113(4):766–786, 2006.

[22] A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognit Psychol*, 12(1):97–136, 1980.

[23] T. Urruty, S. Lew, N. Ihadaddene, and D. A. Simovici. Detecting eye fixations by projection clustering. *ACM Trans. Multimedia Comput. Commun. Appl.*, 3(4):1–20, 2007.

[24] R. Valenti and T. Gevers. Accurate eye center location and tracking using isophote curvature. In *CVPR*, 2008.

[25] J. van de Weijer, T. Gevers, and A. D. Bagdanov. Boosting color saliency in image feature detection. *PAMI*, 28(1):150–156, 2006.

[26] P. Viola and M. J. Jones. Robust real-time face detection. *IJCV*, 57(2):137–154, 2004.

[27] Z. Zheng, H. Wang, and E. Teoh. Analysis of gray level corner detection. *Patt. Recog. Let.*, 20:149–162, 1999.